A new database generation method combining maximin method and kriging prediction for eddy current testing

Sandor Bilicz, Szabolcs Gyimothy, Marc Lambert

A new database generation method combining maximin method and kriging prediction for eddy-current testing

S. Bilicz [*†], Sz. Gyimóthy [†], M. Lambert [*]
[*] Département de Recherche en Électromagnétisme, Laboratoire des Signaux et Systèmes UMR8506
(CNRS-SUPELEC-Univ Paris-Sud), 91192 Gif-sur-Yvette cedex, France
[†] Budapest University of Technology and Economics, Egry J. u. 18, H-1521 Budapest, Hungary
E-mail: bilicz@evt.bme.hu

***Abstract*: The accurate numerical simulation of the eddy-current testing (ECT) experiments usually requires large computational efforts. To avoid the time-consuming computations, the idea – which is new in the domain of ECT – of using *databases* appeared recently. The database consists of well-chosen pairs of input-output samples of a specified ECT experiment. Once it is built, one just has to retrieve the sought data from the database instead of recoursing to the complicated and expensive-to-run simulation methods. However, the generation of such databases is not a straightforward task. This paper presents a new, kriging-based adaptive methodology which yields to databases optimized to the given problem.**

***Keywords*: eddy-current testing, optimized database, kriging**

I. Introduction

The solution of the *forward problems* related to the eddy-current testing configurations is known to be a complicated task. In spite of the vast of contributions and developments in numerical field computation, basically the simulations still need strong computers and they are quite time consuming. Unfortunately, this fact involves the computational difficulties in the related *inverse problems* as well. Moreover, it is often an end-user who aims to solve an inverse problem (e.g. to detect or characterize a defect). However, the end-user may not be a specialist of the underlying theory, he just needs a reliable, *fast* and *simple* method.

Apparently, the mentioned problems can be overcome by changing the "expensive" forward simulator to a "cheap" surrogate model, to a so-called *emulator*. The working principle of the emulator is not (necessarily) based on the underlying physical phenomenon – this is why they are usually much faster and simpler than the simulators.

In this paper, the database, as a kind of emulation is in the focus. The use of the databases in the domain of ECT first appeared in [1]. The main idea is easy to understand. First, one specifies an ECT problem (given specimen, source, receiver and a set of defects that one aims to find). Then the database is being generated: the forward simulator computes the output values (the measurable data) at well-chosen input values (parameters of the defect). All pairs of input-output samples are stored in the database. Finally, when one needs to determine the output at a given input value, the task is not more than to retrieve the sought data from the database. This retrieval can be any kind of interpolation, from the simplest "nearest neighbour" (NN in the followings) method even to the more sophisticated neural network interpolators – the soul of the methods is the same: not to use the precise but expensive simulator.

The main problem of the database generation is the good choice of input samples. One has to find a reasonable compromise between the number of samples in the database and the precision provided by the emulation. So far, mainly adaptive mesh generation methods have been applied to explore the input-output relationship and to sample sequentially the input space (see, e.g. [2]). Our paper presents a new method based on the *kriging prediction*. After the problem definition, the proposed method is described briefly, along with some hints to the basics of kriging. Finally, an illustrative example is presented.

II. Problem definition

Let us assume that in a well-defined ECT configuration, p parameters characterize the defect. These parameters are included in the *input vector* $\mathbf{x}$, taking place in the input space $\mathbb{X}$. The so-called *forward operator* ($\mathcal{F}$) maps $\mathbb{X}$ to the output space $\mathbb{Y}$, by defining the output $\mathbf{y}$ of the ECT experiment:

$$\mathbf{y} = \mathcal{F}\{\mathbf{x}\}.$$

Usually, the output $\mathbf{y}$ is a vector of coil impedances or magnetic flux densities measured at well-defined receiver locations (e.g. on a surface scan). $\mathcal{F}$ hides all underlying physical phenomena, it makes the ECT experiment a "black-box" – which also means that the presented method is general, no theoretical limitation hampers its application for other kind of problems.

As already mentioned, the basic aim is to replace $\mathcal{F}$ by a cheap emulator. Formally, an approximate forward operator – denoted by $\mathcal{G}$ – is to be found. The database-concept offers a convenient way: let us store pre-calculated *samples* (corresponding input-output values) in a database and use an interpolation method based on these samples to approximate $\mathcal{F}$. The goodness of approximation is formalized via the *error of approximation*:

$$h(\mathbf{x}) = \|\mathcal{F}(\mathbf{x}) - \mathcal{G}_n(\mathbf{x})\|_y$$

where $\|\cdot\|_y$ stands for an appropriate norm on the output space. The index of $\mathcal{G}_n$ refers to the fact that the approximate operator is based on the n samples stored in the database: $(\mathbf{x}_1, \mathbf{y}_1), (\mathbf{x}_2, \mathbf{y}_2), \ldots, (\mathbf{x}_n, \mathbf{y}_n)$. The overall accuracy of a database can be characterized by the maximal (over $\mathbb{X}$) approximation error.

The aimed problem in this paper is the method of *generation of optimized databases*. Optimization means that the method takes into account the interpolator (used after the database is built up) and it tries to explore the input-output relationship in the *specific* ECT problem – in order to achieve high accuracy whereas the number of samples is acceptably low.

III. Sequential database building

According to the proposed method, the database generation is an incremental procedure, samples are added one by one. After an initialization, the algorithm is fully automatic. The basic idea is to insert the new sample to the input point $\mathbf{x}_{n+1}$ where the approximation error (at the current stage, with n samples in the database) is the highest:

$$\mathbf{x}_{n+1} = \arg\max_{\mathbf{x}\in\mathbb{X}}\{h(\mathbf{x})\}. \tag{1}$$

This method is obviously reduces the approximation error and seems to be quite simple. However, in spite of its elegance, (1) is extremely expensive to evaluate since each call of $h(\mathbf{x})$ needs the solution of a forward problem. In the followings, we describe how our idea overcomes this difficulty.

The main novelty of our work is that we found a convenient way to predict $h(\mathbf{x})$, in the case when NN interpolation is used. For the sake of completeness, the NN interpolation method is a 0^{th}-order interpolation of $\mathcal{F}\{\mathbf{x}\}$, based on the stored samples. Formally,

$$\mathcal{G}_n(\mathbf{x}) = \mathcal{F}(\mathbf{x}_k) = \mathbf{y}_k,$$
$$\text{where } \|\mathbf{x}-\mathbf{x}_k\|_x \le \|\mathbf{x}-\mathbf{x}_l\|_x \; \forall l = 1,2,\ldots,n.$$

The symbol $\|\cdot\|_x$ stands for an appropriate norm on $\mathbb{X}$.

The method is based on the concept of the so-called *distance functions*, defined as

$$Q_k(\mathbf{x}) = \|\mathcal{F}(\mathbf{x}) - \mathbf{y}_k\|_y, \text{ where } k = 1,2,\ldots,n.$$

Obviously, for a given $\mathbf{x}$ input, the $h(\mathbf{x})$ approximation error (using NN interpolation) is equal to $Q_{NN}(\mathbf{x})$, the distance function related to the nearest neighbour of $\mathbf{x}$. However, the distance functions are expensive-to-evaluate since they need the solution of a forward problem. Thus, they cannot be used directly to express $h(\mathbf{x})$. The key point of our method is here: we build a *kriging model* (based on the n samples already added to the database) for their prediction, instead of the precise evaluation.

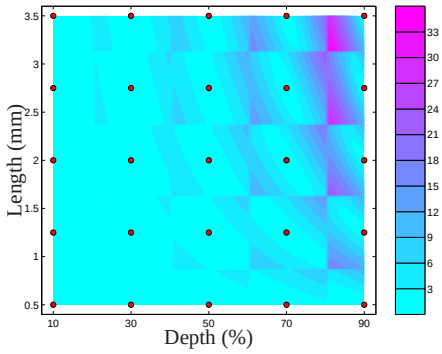
Kriging provides an elegant prediction for all the n distance functions all over $\mathbb{X}$, thus, the approximation error can be predicted as well, making (1) easy to evaluate approximately.

Kriging is a stochastic tool for predicting unknown functions, based on some observations (by now, kriging is a well-known and widely used method in several domains, see e.g. [3]). Its main idea is to model the function by a Gaussian process. Based on the observations, the mean of the process is predicted, which passes through all observation points, thus it is an interpolation method. The predicted mean is expressed as a linear combination of the observations.

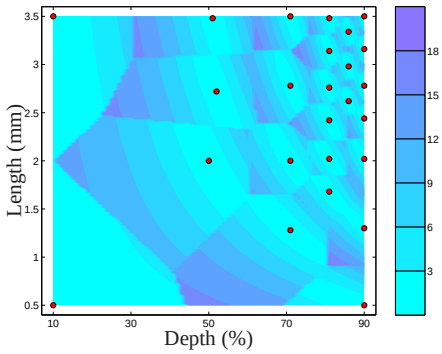
By repeating these approximation error prediction – sample insertion cycles till a stopping criterion is passed, one gets the algorithm for building optimized databases. The stopping criterion can be related for instance to the predicted maximal $h(\mathbf{x})$ and/or to the predicted uncertainty of the kriging interpolation.

IV. Illustrative example

A simple example with a 2 dimensional input space is presented here. A thin OD type, cuboid-shaped defect is characterized by its length and depth. In Fig. 1 a comparison is shown between the performances of a "classical" database (regular sampling in $\mathbb{X}$) and an optimized one (made by using the presented kriging-based method). The approximation error is normalized by the norm of the impedance signal of the largest possible defect (3.5 mm, 90 %) and given in % by the color map. Both databases consists of 25 samples. One can see that the optimized database outperforms the classical one in accuracy.



(a) Regularly sampled input space.



(b) Optimized database.

Figure 1. Normalized approximation error (color map, %) and input samples (red markers).

V. Conclusions and further development

A new, kriging-based method for optimized database generation is presented. In the light of the preliminary results, the methodology seems to be promising. Theoretically, there is no limitation for the number of defect parameters.

In this abstract, only the simplest approach has been presented yet, in the authors' opinion, kriging offers further possibilities to develop more sophisticated methods for database generation. Note that also the uncertainty of the prediction can be expressed in a closed form – this uncertainty might be useful to improve the performance of the algorithm.

References

[1] J. Pávó and S. Gyimóthy, "Adaptive inversion database for electromagnetic nondestructive evaluation," *NDT&E International*, vol. 40, pp. 192–202, 2007.

[2] S. Gyimóthy, I. Kiss, and J. Pávó, "Adaptive sampling technique based on moving meshes for building data-equidistant inversion databases for NDT," in *OIPE 2008*, Ilmenau (Germany), September 2008, pp. 139–140.

[3] J. Chiles and P. Delfiner, *Geostatistics, Modeling Spatial Uncertainty*. Wiley, 1991.